Coalitional Ability in Multi-Agent Systems: A Logical Approach

Marc Pauly
Center for Mathematics and Computer Science (CWI)
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
E-mail: pauly@cwi.nl

Abstract

Effectivity frames are introduced as a model of ability in multi-agent systems, i.e. of what groups of agents can achieve by coordinated action in dynamic processes such as extensive games with or without simultaneous moves. Local effectivity is distinguished from different kinds of global and terminal effectivity, respectively what groups of players can maintain throughout and what they can achieve eventually. We provide an example of how effectivity frames can be used to formalize (1) questions about goal achievement and maintenance, (2) the synthesis problem for multi-agent systems, and (3) the interplay between local and global properties of such systems. A modal logic for local and global coalitional effectivity is presented and we show how the three problems mentioned can be translated into this logical framework and solved by standard logical methods.

Introduction

Modeling actions and their effects is a task which has occupied many researchers in computer science, logic, economics and arti cial intelligence. In the simplest case, we have one agent (person, process) who can choose between taking different actions which change the state of the world in various ways. A simple model of this scenario will contain an accessibility relation R which associates to every state of the world all those states which the agent can bring about through his actions, i.e. sRt holds if the agent can act in state s as to bring about state t. In modal logic, one introduces a language to talk about such *Kripke* models: $\Diamond\varphi$ expresses that the agent can act in such a way that φ will be true after his action.

When generalized to multiple agents, Kripke models do not suf ce anymore as a model of ability. Instead we will employ effectivity functions (Moulin & Peleg 1982; Abdou & Keiding 1991) which associate to every state and every group of agents C the sets of states for which C is effective. Since the resulting states can again be associated with an effectivity function, we obtain a dynamic model of what groups of agents can achieve by joint action. At each state of the model, we can distinguish local effectivity from different kinds of terminal effectivity: While a group of agents may not be able to bring about φ immediately, the group may be able to achieve φ eventually. As we will show, extensive games, nondeterministic processes and voting procedures can all be viewed as instances of these general *effectivity frames*. We then present some applications of the model by means of a simple voting scenario. Various questions about goal achievement, the synthesis of multi-agent systems etc. will be formalized using effectivity frames. To describe these frames, we use a modal logic in which properties of effectivity frames can be expressed as modal axioms, and questions e.g. about the existence of a particular system implementing a given multi-agent speci cation can be expressed as satis ability problems. While we do not study the logic itself in the present paper, comments on some of its meta-theoretic properties are made.

A Dynamic Model of Effectivity

Effectivity Functions

Throughout this paper, we assume that a nonempty nite set N of agents or players is given, as well as a nonempty set of states S. An *effectivity function* $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ associates to every group of players the sets of outcomes for which the group is effective. For a coalition $C \subseteq N$, $X \in E(C)$ will hold iff the players in C have a joint strategy for bringing about an outcome in X. We shall require that effectivity functions are monotonic, i.e. $X \subseteq Y \subseteq S$ implies that for every coalition $C \subseteq N$, if $X \in E(C)$ then $Y \in E(C)$. Given the intuitions put forth, this requirement should be a natural one.

We present some further properties of effectivity functions which will play a role later: In many circumstances one will want to assume that a group which becomes larger has possibly more power but certainly not less. In that case, E is *coalition-monotonic*, i.e. for $C \subseteq C' \subseteq N$, $E(C) \subseteq E(C')$. As a basic consistency requirement, we usually want to exclude cases where complementary coalitions are effective for complementary things, for in that case, both coalitions could use their power and end up in an inconsistent situation. The notion of regularity captures this concern: E is *C-regular* if for all X, if $X \in E(C)$ then $\overline{X} \notin E(\overline{C})$. As a converse to regularity, call E *C-maximal* if for all X, if $\overline{X} \notin E(\overline{C})$ then $X \in E(C)$. E is *regular* (*maximal*) iff for all coalitions C it is C-regular (C-maximal). If we think of a two-player game with two possible outcomes win_1 and

win_2, regularity expresses that the game is zero-sum: If player 1 wins ($\{1\}$ is effective for $\{win_1\}$) then player 2 cannot win ($\{2\}$ is not effective for $\{win_2\}$). Similarly, maximality expresses determinacy: If player 1 does not win ($\{1\}$ is not effective for $\{win_1\}$) then player 2 must win ($\{2\}$ is effective for $\{win_2\}$). Finally, the most interesting principle governs the formation of coalitions. It states that coalitions can combine their strategies to (possibly) achieve more: E is *superadditive* if for all X_1, X_2, C_1, C_2 such that $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$ imply that $X_1 \cap X_2 \in E(C_1 \cup C_2)$.

Effectivity Frames & Structures

The dynamic models of effectivity are captured by the notion of an *effectivity frame* $\mathcal{F} = (S, E)$, where S is the nonempty set of states and $E : \mathcal{S} \to (\mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S)))$ is the *local effectivity structure* which associates to every state $s \in S$ an effectivity function $E(s)$, i.e. $X \in E(s)(C)$ if at state s, coalition C is effective for the set of states X. Given that effectivity functions are linked to states and that the alternatives which can be brought about are states again, frames do present a dynamic model of effectivity. For easier readability, we shall often write sE_CX instead of $X \in E(s)(C)$, and we will sometimes view the effectivity structure E as a collection of maps $E_C : \mathcal{P}(S) \to \mathcal{P}(S)$ with $E_C(X) = \{s \in S | sE_CX\}$. Monotonicity can now be formulated as the requirement that if $X \subseteq X'$ then $E_C(X) \subseteq E_C(X')$. A state s where no coalition is effective for anything is called a *terminal state*, i.e. s is terminal iff there is no coalition C and no set of states X such that sE_CX. We denote the set of terminal states $\{s \in S | \neg\exists C \exists Y sE_CY\}$ as $S_\perp$.

Lastly, note that all of the properties of effectivity functions which we introduced previously can be lifted to frames/structures: A frame (S, E) has a given property (e.g. maximality) iff every $E(s)$ has it.

Global and Terminal Effectivity

Given an effectivity frame $\mathcal{F} = (S, E)$ which contains information about effectivity at every state, we can also investigate effectiveness in the long run. For this, we de ne the *global effectivity structure* $E^* : \mathcal{S} \to (\mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S)))$, where a coalition C is globally effective for X if it can maintain X throughout the game/process. Again, the global effectivity structure determines a (global) effectivity function $E^*(s)$ for every state $s \in S$. Viewing E^*_C as a function from $\mathcal{P}(S)$ to $\mathcal{P}(S)$, we de ne

$$E^*_C(Z) = \bigcup\{X \subseteq S | X \subseteq Z \cap (S_\perp \cup E_C(X))\}$$

By the Knaster-Tarski theorem, since E_C is a monotonic operator, $E^*_C(Z)$ is the greatest xpoint of the operator $F^Z_C(X) = Z \cap (S_\perp \cup E_C(X))$, i.e. $F^Z_C(E^*_C(Z)) = E^*_C(Z)$ and for every $A \subseteq S$ such that $F^Z_C(A) = A$ we have $A \subseteq E^*_C(Z)$. Using ν-notation, we can write

$$E^*_C(Z) = \nu X.F^Z_C(X) = \nu X.Z \cap (S_\perp \cup E_C(X))$$

Intuitively, C is globally effective for X at state s if $s \in X$ and C can subsequently maintain X, i.e. if s is not a terminal state, C is locally effective for X.

Often we are not interested in global effectivity generally but in *terminal effectivity*, the terminal states or nal outcomes which a coalition can bring about. Analogous to program correctness two notions of effectivity can be distinguished: A coalition C is *partially effective* for X if it can guarantee X to hold at every terminating (i.e. nite) run of the game/process. Thus, the strategy for C may still lead to an in nite run. In contrast, C is *totally effective* for X if it is partially effective for X and can guarantee termination at the same time. Formally, de ne $E^p, E^t : \mathcal{S} \to (\mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S)))$ as the partial and total effectivity structures respectively as follows (where μX denotes the least xpoint):

$$E^p_C(X) = E^*_C(\overline{S_\perp} \cup X) = \nu Y.(\overline{S_\perp} \cup X) \cap (S_\perp \cup E_C(Y))$$
$$E^t_C(X) = \mu Y.(S_\perp \cap X) \cup E_C(Y)$$

For total effectivity, we start with the terminal states which are in X and iteratively add to them the states where C can guarantee an outcome among those states. For partial effectivity, we require that C is globally effective for $\overline{S_\perp} \cup X$, i.e. C can maintain forcing states which are either nonterminal or in X.

While effectivity frames do not formally distinguish a special initial state s_0, in the modelling applications we shall often assume such a distinguished initial state; consequently, e.g. the total effectivity function associated to such a frame will denote the total effectivity function of the initial state of the frame.

The following results capture the main relationships between global, partial and total effectivity. These results should be seen as generalizations of Dijkstra's work in (Dijkstra 1976) which discusses the links between partial and total program correctness.

Proposition 1

1. *For every effectivity frame $\mathcal{F} = (S, E)$, total effectivity implies partial effectivity, i.e. $E^t_C(X) \subseteq E^p_C(X)$.*
2. *For every superadditive effectivity frame $\mathcal{F} = (S, E)$, at states where no in nite play is possible, partial effectivity implies total effectivity, i.e. $E^t_\emptyset(S) \cap E^p_C(X) \subseteq E^t_C(X)$.*
3. *For every C-regular and C-maximal effectivity frame $\mathcal{F} = (S, E)$, total effectivity is the dual of partial effectivity, i.e. $\overline{E^t_C(X)} = E^p_{\overline{C}}(\overline{X}) = E^*_{\overline{C}}(\overline{S_\perp \cap X})$.*

Special Effectivity Frames: Games

Extensive Games with Simultaneous Moves

Effectivity frames generalize many dynamic game models used in game theory. In an extensive game form with simultaneous moves (Osborne & Rubinstein 1994), players may act simultaneously at every stage of the game in order to determine the resulting game position. Formally, to obtain such a model, we associate a strategic game with every state of the world. A *strategic game* $G = (N, \{\Sigma_i | i \in N\}, o, S)$ consists of the set of agents N, a nonempty set of strategies or actions Σ_i for every player $i \in N$, the set of states S and an outcome function $o : \Pi_{i \in N}\Sigma_i \to S$ which associates with every tuple of strategies of the players (strategy pro le) an outcome state in S.

In game theory (Osborne & Rubinstein 1994; Binmore 1992), strategic games also come equipped with a preference relation $\succeq_i \subseteq S \times S$ for every player $i \in N$ which indicates which outcomes a player prefers. Strictly speaking, our strategic games are only game *forms* which can be turned into a game by adding these preference relations.

For notational convenience, let $\sigma_C := (\sigma_i)_{i \in C}$ denote the strategy tuple for coalition $C \subseteq N$ which consists of player i choosing strategy $\sigma_i \in \Sigma_i$. Then given two strategy tuples σ_C and $\sigma_{\overline{C}}$ (where $\overline{C} := N \setminus C$), $o(\sigma_C, \sigma_{\overline{C}})$ denotes the outcome state associated with the strategy pro le induced by σ_C and $\sigma_{\overline{C}}$.

Let Γ_S^N be the set of all strategic games between the set of players N over the set of states S. Then we de ne an *extensive game (with simultaneous moves)* as a partial function $\gamma : S \hookrightarrow \Gamma_S^N$ which associates strategic games to non-terminal states.

The notion of effectivity which we associate with strategic games is known as α-effectivity, (Moulin & Peleg 1982; Moulin 1983; Abdou & Keiding 1991). Given a game G, a coalition $C \subseteq N$ will be *α-effective* for a set $X \subseteq S$ iff the coalition has a joint strategy which will result in an outcome in X no matter what strategies the other players choose. Formally, the *α-effectivity function* $E_G^\alpha : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ of a strategic game G is de ned as

$$X \in E_G^\alpha(C) \text{ iff } \exists \sigma_C \forall \sigma_{\overline{C}} \; o(\sigma_C, \sigma_{\overline{C}}) \in X$$

We say that an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *α-corresponds* to a strategic game G iff $E = E_G^\alpha$. Similarly, an effectivity frame $\mathcal{F} = (S, E)$ *α-corresponds* to an extensive game $\gamma : S \hookrightarrow \Gamma_S^N$ provided that $sE_C X$ iff $\gamma(s)$ is de ned and $X \in E^{\alpha}_{\gamma(s)}$.

The question to be examined now is which effectivity frames α-correspond to some extensive game. Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *playable* iff it satis es the following 4 conditions: (1) $\forall C \subseteq N : \emptyset \notin E(C)$, (2) $\forall C \subseteq N : S \in E(C)$, (3) E is N-maximal, and (4) E is superadditive. Proposition 2 below extends characterization results obtained in (Moulin 1983; Peleg 1998). A proof can be found in (Pauly 2000).

Proposition 2 *An effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ α-corresponds to a strategic game $G \in \Gamma_S^N$ iff E is playable.*

Corollary 3 *An effectivity frame $\mathcal{F} = (S, E)$ α-corresponds to an extensive game $\gamma : S \hookrightarrow \Gamma_S^N$ iff for every state $s \notin S_\perp$, $E(s)$ is playable.*

We shall sometimes refer to an effectivity frame $\mathcal{F}$ simply as an extensive game if it α-corresponds to such a game.

Claims (1) and (2) of proposition 1 apply to extensive games, whereas claim (3) does not, since extensive games are not necessarily maximal. Consider e.g. the extensive game with simultaneous moves $(\{s_0, s_1, s_2, s_3, s_4\}, \gamma)$ where γ is de ned only for s_0 yielding the strategic game depicted below.

Consider the effectivity frame $\mathcal{F} = (S, E)$ which α-corresponds to this extensive game. Note that $E(s_0)$ is not maximal since the row player is not effective for $\{s_1, s_4\}$

	left	right
up	s_1	s_2
down	s_3	s_4

Figure 1: A non-maximal game where partial and total effectivity are not duals.

while the column player is not effective for $\{s_0, s_2, s_3\}$ either. Furthermore, while the row player is not totally effective for $\{s_1, s_4\}$ at s_0, the column player is not partially effective for $\{s_0, s_2, s_3\}$ either.

Extensive Games without Simultaneous Moves

A special case of the extensive games discussed in the previous section arises if at every stage of the game, one of the players is in complete control in determining the next stage. In such an extensive game without simultaneous moves, every stage of the game has a local dictator. Formally, we call a strategic game $G = (N, \{\Sigma_i | i \in N\}, o, S)$ a *dictatorship* iff there is some $d \in N$ such that

$$\forall \sigma_d \; \exists s \; \forall \sigma_{N \setminus \{d\}} \;\; o(\sigma) = s$$

In such a dictatorship, there is an individual d (the dictator) whose choices completely determine the outcome state, independent of what the others do. Note that in case there is more than one dictator, the outcome function is constant (i.e. $\exists s \forall \sigma \; o(\sigma) = s$) and hence every player is a dictator.

Let Δ_S^N be the set of all dictatorships for the set of players N over the set of states S. We de ne an *extensive game without simultaneous moves* as a partial function $\gamma : S \hookrightarrow \Delta_S^N$ which associates dictatorships to non-terminal states. Note that given an initial state, we can picture an extensive game without simultaneous moves as a standard game tree where nodes correspond to states which are labeled with the local dictator of that state, i.e. the player who is to move at that state (see e.g. gure 4). Note that in case $|N| = 1$ or in case the local dictator is the same at every state, the game is a one-player game, or equivalently, a process.

Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *individualistic* iff it is playable and $E(N) = \bigcup_{i \in N} E(\{i\})$. The condition ensures that everything which can be forced can be forced already by some individual. The following result (proved in (Pauly 2000)) shows that individualism is an extremely strong assumption: While it seems to say only that the whole is equal to the sum of its parts, due to superadditivity, it actually says that the whole is equal to one particular part.

Proposition 4 *An effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ α-corresponds to a dictatorship $G \in \Delta_S^N$ iff E is individualistic.*

Put positively, unless we have a dictatorship, coalitions of agents can sometimes achieve more than their members individually, cooperation is thus advantageous.

Corollary 5 *An effectivity frame $\mathcal{F} = (S, E)$ α-corresponds to an extensive game without simultaneous moves $\gamma : S \hookrightarrow \Delta_S^N$ iff for every state $s \notin S_\perp$, $E(s)$ is individualistic.*

Since individualistic effectivity functions are regular and maximal, all three claims of proposition 1 apply to extensive games without simultaneous moves.

Democracies

Effectivity frames can also be utilized to model the power of coalitions in voting procedures. Consider a strategic game $G = (N, \{\Sigma_i | i \in N\}, o, S)$ where $|N|$ is odd, $\Sigma_i = \{yes, no\}$ for all $i \in N$ and there are states $s_y, s_n \in S$ (we allow for $s_y = s_n$) such that for every strategy pro le σ, $o(\sigma) \in \{s_y, s_n\}$ and $o(\sigma) = s_y$ if $|\{\sigma_i | \sigma_i = yes\}| > \frac{1}{2}|N|$ and $o(\sigma) = s_n$ otherwise. This strategic game corresponds to a vote between two alternatives where each member of N participates and the outcome state is determined by the majority of the votes. We shall call such a voting game a *2-alternative majority vote*. Since an even number of voters creates the problem of how to resolve ties, we require $|N|$ to be odd here. Nonetheless, this simple example should demonstrate how results for more elaborate "democratic" voting procedures (e.g. including a distinguished chairman who decides in case of a tie) can be obtained.

Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *majorative* iff for every coalition C with $|C| > \frac{1}{2}|N|$ we have $E(N) \subseteq E(C)$. E is *binary* iff

$$X \in E(N) \text{ and } \overline{X} \cap Y \in E(N) \Rightarrow X \cup Y \in E(\emptyset)$$

Quite naturally, the rst condition formalizes that a majority suf ces to establish anything, and the second condition captures that there are only two alternatives to be chosen from. Note that if E is playable and majorative, E is also maximal and furthermore $\overline{X} \in E(N) \Rightarrow X \notin E(C)$ for $|C| < \frac{1}{2}|N|$.

Proposition 6 *An effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ α-corresponds to a 2-alternative majority vote $G = (N, \{\Sigma_i | i \in N\}, o, S)$ iff E is playable, majorative and binary.*

De ning a *democratic binary procedure* as a partial function which associates 2-alternative majority votes to non-terminal states, we obtain

Corollary 7 *An effectivity frame $\mathcal{F} = (S, E)$ α-corresponds to a democratic binary procedure iff for every state $s \notin S_\perp$, $E(s)$ is playable, majorative and binary.*

Since playable majorative effectivity functions are also maximal, proposition 1 applies to democracies as well.

Some Applications

Consider a political body $N = \{1, 2, 3, 4, 5, 6\}$ which has to decide on passing a new law. First, a subcommittee $D = \{2, 3, 4\}$ has to decide (by majority) which precise version of the law is to be presented to the full political body. Subsequently, the whole political body decides whether the law is passed or not. Again, the majority of the votes decides, and in case of a draw, the vote of the chairman 1 is decisive. If the law (as proposed by committee D) is not passed, the initiative is returned to committee D which has to make a new proposal for the law, and the process repeats itself.

We assume for simplicity that there are only two versions of the law which are under discussion, version 1 and version 2. If the body N rejects the proposal of committee D, the committee can either decide to propose the other version of the law, or it can resubmit its original proposal, possibly resulting in a stalemate which may turn into an in nite loop (some might claim that this model is suf ciently realistic to capture the essentials of the legislative process in some countries). Figure 2 depicts the situation as a graph.

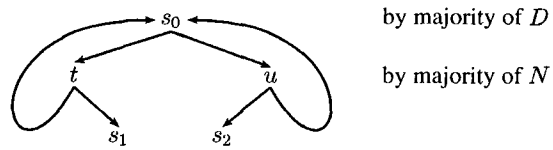


Figure 2: An example of binary majority voting with subcommittees

One can think of the situation described in terms of coalitional effectivity: $s_0 E_C X$ holds iff at state s_0, coalition C can force the local voting outcome to lie in set X, i.e. iff one of the following two conditions is met: (1) $\{t, u\} \cap X \neq \emptyset$ and $|C \cap D| > 1$, or (2) $\{t, u\} \subseteq X$. Analogous de nitions can be given for $tE_C X$ and $uE_C X$, incorporating the special role of the chairman.

We can also use this example to illustrate global, partial and total effectivity: Figure 3 displays some interesting examples which demonstrate the unequal powers of four 3-player coalitions at the initial state s_0.

coalition/states	$\{s_1\}$	$\{s_1, s_2\}$	$\{s_0, t, u\}$
$\{1, 2, 3\}$	t	t	g
$\{2, 3, 4\}$	p	p	–
$\{1, 4, 5\}$	–	t	g
$\{4, 5, 6\}$	–	–	–

Figure 3: Global (g), total (t) and partial (p) effectivity in the voting example of gure 2 at state s_0.

This very simple toy example will serve to illustrate a number of different problems which can be raised and formalized using effectivity frames.

Strategy-Proofness

As the designer of a voting procedure such as gure 2, we may want to know whether it can be manipulated in ways we consider undesirable, i.e. whether an agent or a group of agents has a strategy to achieve an outcome which it considers desirable but which we as the designer of the procedure would consider undesirable in terms of the social welfare of all agents. For example, a certain group of agents may have an incentive to delay passing a new law as long as possible, preferably inde nitely. As the designer, we may want to design our voting procedure to prevent any group of agents from steering the voting process into an in nite loop. As it turns out, the voting procedure of gure 2 is not strategy-proof in this respect: Both $\{1, 2, 3\}$ and $\{1, 4, 5\}$ can glob-

ally maintain $\{s_0, t, u\}$, i.e. they have the power to keep the process going forever, never reaching any decision.

Goal Achievement/Maintenance Problems

The typical questions considered in the literature on agent cooperation and agent design concern goal achievement and goal maintenance (Tennenholtz & Moses 1989; Wooldridge 2000): Given a particular multi-agent system, is there a run of the system in which all agents achieve their respective goals? Translated into our framework, given n agents with goals G_1 through G_n, the formula $s_0 \notin E^*_\emptyset(\overline{G_1 \cap \ldots \cap G_n})$ expresses that not every run of the system violates all goals inde nitely, i.e. there is a run of the system where all the goals will eventually be simultaneously satis ed. Similarly, $s_0 \in E^t_N(G_1 \cap \ldots \cap G_n)$ expresses that there is a run of the system where all goals are satis ed at the end. In our voting example, the agents can cooperatively achieve the passing of any law, a minimal positive requirement for any voting system it would seem: $s_0 \in E^t_N(\{s_1\})$ and $s_0 \in E^t_N(\{s_2\})$ both hold.

Similarly, we can inquire about the feasibility of a particular set of outcomes for a single agent or a group of agents generally. The coalition $\{1, 4, 5\}$ cannot force a particular version of the law to be passed, whereas $\{2, 3, 4\}$ does have at least partial power to do so: If some mechanism in the legislative process would rule out an in nite stalemate, this coalition can guarantee any outcome; still, it is unable to rule out such an in nite stalemate on its own, i.e. it is partially but not totally effective for $\{s_1\}$ as well as for $\{s_2\}$. This distinction between partial and total goal achievement based on the possibility of in nite runs does not seem to have received much attention in the literature on agent cooperation.

Finally, instead of achieving a particular goal, one may want to know whether a group of agent's can maintain a particular state of affairs. As mentioned previously, both $\{1, 2, 3\}$ and $\{1, 4, 5\}$ can maintain the set $\{s_0, t, u\}$, i.e. the state of affairs where no law has been passed yet, $s_0 \in E^*_{\{1,2,3\}}(\{s_0, t, u\}) \cap E^*_{\{1,4,5\}}(\{s_0, t, u\})$.

Comparing Coalitional Power

Based on the different global, partial and terminal abilities of the various coalitions, one can obtain an ordering of groups of agents with respect to their abilities. Inspecting gure 3, the following partial order of these four 3-player coalitions emerges:

$$\{4, 5, 6\} \prec \{2, 3, 4\}, \{1, 4, 5\} \prec \{1, 2, 3\}$$

Note that the coalitions $\{2, 3, 4\}$ and $\{1, 4, 5\}$ are incomparable: On the one hand, coalition $\{2, 3, 4\}$ is more powerful since it is able to pass any law provided that the procedure terminates eventually. On the other hand, coalition $\{1, 4, 5\}$ can force the procedure to terminate or keep it going forever, something which coalition $\{2, 3, 4\}$ cannot do.

Multi-Agent Synthesis

The question whether given some environment, there exists an agent who can achieve a certain goal or maintain a certain state of affairs has been considered e.g. in (Wooldridge 2000). An analogous question can be asked when there are multiple agents: Is there a multi-agent system in which certain groups of agents can achieve some given goals and maintain certain states of affairs? Formally speaking, the global and terminal effectivity functions are partially spec- i ed for particular groups of agents, and we want to know whether there is an effectivity frame whose derived global and terminal effectivity functions satisfy those properties. Ideally, we would want automatic multi-agent synthesis which not only answers this question but provides us with an implementation or realization of the speci cation. Thus, whereas we previously considered whether a given effectivity frame such as the voting procedure of gure 2 satis- es some multi-agent speci cations of achievement, maintenance, etc., we now would like to obtain such a procedure from some speci cation directly.

The multi-agent synthesis problem also has a more abstract game-theoretic version. Given a particular (possibly only partially speci ed) effectivity function, one may ask whether it can be implemented or realized by means of a particular procedure. An agenda setter might be interested to nd out whether a given power distribution can be realized by a (democratic) voting procedure of a particular kind. More abstractly, given a strategic game G with its associated α-effectivity function E^α_G, one might want to know whether there is an extensive game without simultaneous moves G' such that $E^t_{G'} = E^\alpha_G$.

As a very simple example of a realization problem using α-effectivity, consider again the strategic game form G of gure 1. Since there is no extensive game G' without simultaneous moves such that $E^t_{G'} = E^\alpha_G$, there can also be no extensive game with a reduced strategic form equivalent to G. In contrast, if we replace s_4 by s_3, the extensive game G' of gure 4 satis es $E^t_{G'} = E^\alpha_G$ (player 1 plays rows, player 2 columns) and has an equivalent reduced strategic form.

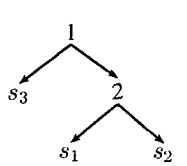
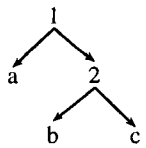
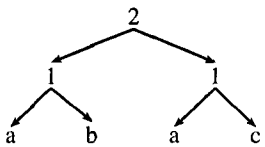


Figure 4: An extensive game without simultaneous moves

The question, whether the α-effectivity function of a strategic game can be realized by an extensive game, is closely related to a question that has received some attention in the game theory literature (see e.g. (Abdou 1998)): For which strategic game forms does there exist an extensive game form (of perfect information, without simultaneous moves) with the same *reduced strategic form*? Two strategic games with the same reduced strategic form have the same α-effectivity function while the converse does not hold, as the following two extensive games show:



1\2	l	r
l	a	a
r	b	c



1\2	l	r
ll	a	a
lr	a	c
rl	b	a
rr	b	c

The table to the right of each game represents its (reduced) strategic form. From the perspective of α-effectivity, both extensive games are the same, whereas the strategic form of the game on the right reveals that player 1 has strategic options unavailable to him in the game on the left: In the right game, player 1 has a strategy lr which guarantees him either outcome a or outcome c. From the point of view of α-effectivity, this strategy is negligible given that he has a "stronger" strategy ll which guarantees outcome a. Still, player 1 may prefer strategy lr which forces $\{a, c\}$ to strategy ll which forces $\{a\}$ if he strictly prefers c to a. So from a perspective which includes preferences, one may not want to identify the two games.

Local vs. Global Properties

Besides the questions raised previously which were of a rather concrete nature, effectivity frames can also be utilized to study some more abstract questions. The different kinds of effectivity frames associated e.g. with extensive games and democracies have been de ned in terms of local requirements, i.e. properties which the local effectivity functions had to satisfy. Some of these properties will be maintained globally or terminally, some will not. One can show e.g. that for games without in nite plays, the total (= partial) effectivity function is playable:

Proposition 8 *If $\mathcal{F} = (S, E)$ α-corresponds to an extensive game and $s \in E^t_\emptyset(S)$, then $E^t(s)$ is playable.*

Consequently, for every extensive game G without in nite plays, there is a strategic game G' such that the total effectivity function of G is the α-effectivity function of G'. In fact, one such strategic game G' is simply the strategic normal form of G (see (Osborne & Rubinstein 1994)). Note also that the global effectivity function E^* is not playable: In the simple 1-player game where at the initial state s, the player has only one possible move resulting in a nal state t, E^* is not N-maximal: neither $sE^*_\emptyset(\overline{S_\perp})$ nor $sE^*_N(S_\perp)$ holds.

Considering democracy as a further example, while the binary aspect of a democratic binary procedure may be lost globally, the democratic aspect is maintained: If E is majorative at every state of an effectivity frame, then so is E^t, so a democratic procedure will maintain democracy overall. That the converse is not true can be gathered from the extensive game in gure 5 where at the initial state, E^t is majorative while E is not.

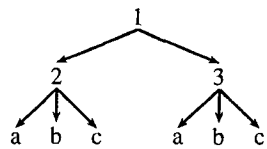


Figure 5: A game which is majorative terminally but not locally.

A Logical Framework

Syntax and Semantics

Coalition Logic, introduced in (Pauly 2000), provides a formal system to reason about effectivity frames. The formula $[C]\varphi$ is true at a state provided that coalition C is locally effective for achieving a state where φ holds. In this paper, we extend the purely modal system of (Pauly 2000) by adding a new operator for global effectivity: $[C^*]\varphi$ is true at a state if coalition C is globally effective for φ. The resulting logic can be viewed as a generalized multi-player version of the game logic proposed in (Parikh 1985).

Given the set of agents N, we de ne the syntax of Coalition Logic as follows. Given a set of atomic propositions Φ_0, a formula φ can have the following syntactic form:

$$\varphi := \quad \perp \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid [C]\varphi \mid [C^*]\varphi$$

where $p \in \Phi_0$ and $C \subseteq N$. We de ne $\top$, $\wedge$, $\rightarrow$ and $\leftrightarrow$ as usual: $\top := \neg\perp$, $\varphi \wedge \psi := \neg(\neg\varphi \vee \neg\psi)$, $\varphi \rightarrow \psi := \neg\varphi \vee \psi$ and $\varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. In case $C = \{i\}$, we we write $[i]\varphi$ instead of $[\{i\}]\varphi$. Furthermore, we de ne terminal effectivity in terms of global effectivity as is to be expected given the semantic de nitions of E^t and E^p, i.e. we use the following abbreviations:

$$\begin{aligned} \sigma_\perp &:= \bigwedge_{C \subseteq N} \neg[C]\top \\ [C^p]\varphi &:= [C^*](\varphi \vee \neg\sigma_\perp) \\ [C^t]\varphi &:= \neg[\overline{C}^*]\neg(\varphi \wedge \sigma_\perp) \end{aligned}$$

Note that the de nition of $[C^t]\varphi$ relies on proposition 1 which applies only to regular and maximal frames, but since all of the frames we will be dealing with in our applications fall into that category, we can cut back on the number of primitive operators.

An *effectivity model* $\mathcal{M} = (S, E, V)$ consists of an effectivity frame $\mathcal{F} = (S, E)$ and a valuation for the propositional letters $V : \Phi_0 \rightarrow \mathcal{P}(S)$. Given such a model, truth of a formula in a model at a state is de ned as follows

$$\begin{array}{lll} \mathcal{M}, s \not\models \perp & & \\ \mathcal{M}, s \models p & \text{iff} & p \in \Phi_0 \text{ and } s \in V(p) \\ \mathcal{M}, s \models \neg\varphi & \text{iff} & \mathcal{M}, s \not\models \varphi \\ \mathcal{M}, s \models \varphi \vee \psi & \text{iff} & \mathcal{M}, s \models \varphi \text{ or } \mathcal{M}, s \models \psi \\ \mathcal{M}, s \models [C]\varphi & \text{iff} & sE_C\varphi^{\mathcal{M}} \\ \mathcal{M}, s \models [C^*]\varphi & \text{iff} & sE^*_C\varphi^{\mathcal{M}} \end{array}$$

where $\varphi^{\mathcal{M}} = \{s \in S | \mathcal{M}, s \models \varphi\}$. A formula φ is *valid in* $\mathcal{M}$ iff $\varphi^{\mathcal{M}} = S$, and φ is valid iff φ is valid in all effectivity models. A set of formulas Ψ is *satis able* iff there is a model $\mathcal{M}$ for which $\bigcap_{\psi \in \Psi} \psi^{\mathcal{M}} \neq \emptyset$.

Applications

The purpose of this section is to show how the applications mentioned before can all be translated into logical questions of model checking, validity and satis ability.

Reconsidering rst the voting procedure of gure 2, one can easily translate the facts about global and terminal effectivity into our logical language: Let $\mathcal{M} = (S, E, V)$ be the model which captures the procedure depicted the gure, where $\Phi_0 = \{p_1, p_2, q\}$ and $V(p_i) = \{s_i\}$, and $V(q) = \{s_0, t, u\}$. Questions of goal achievement, maintenance and strategy-proofness are now simple model checking questions for coalition logic formulas. At the initial state, $\mathcal{M}, s_0 \models [\{1,2,3\}^t]p_1 \wedge [\{1,2,3\}^t]p_2 \wedge [\{1,2,3\}^*]q$, i.e. the coalition $\{1,2,3\}$ can achieve any possible outcome as well as a stalemate, whereas the coalition $\{1,4,5\}$ is weaker, $\mathcal{M}, s_0 \models [\{1,4,5\}^t](p_1 \vee p_2)$ but $\mathcal{M}, s_0 \not\models [\{1,4,5\}^t]p_1 \vee [\{1,4,5\}^t]p_2$. Furthermore, $\mathcal{M}, s_0 \models [\{1,4,5\}^*]q$, so this coalition can block any law from getting passed. Even weaker, coalition $\{4,5,6\}$ has virtually no power, since its counter-coalition $\{1,2,3\}$ is all-powerful. Thus, these facts about truth in a given model are the logical analogue of gure 3.

Whereas goal achievement and maintenance questions were formulated in terms of model checking, multi-agent synthesis can be turned into a satis ability problem. Consider again the realization problem discussed in the section on applications which we will formulate axiomatically. We examine again the strategic game G of gure 1 with $N = \{1,2\}$ and its associated α-effectivity function E^α_G. We use $\Phi_0 = \{a, b, c, d\}$ for the outcomes of the game and specify that the outcomes are mutually exclusive, complete and hold only at terminal states:

$$
\begin{array}{l}
[\emptyset^*]\ (a \to \neg b \wedge \neg c \wedge \neg d) \\
[\emptyset^*]\ (b \to \neg a \wedge \neg c \wedge \neg d) \\
[\emptyset^*]\ (c \to \neg a \wedge \neg b \wedge \neg d) \\
[\emptyset^*]\ (d \to \neg a \wedge \neg b \wedge \neg c) \\
[\emptyset^*]\ (\sigma_\perp \leftrightarrow (a \vee b \vee c \vee d))
\end{array}
$$

Next, we require that all outcomes are possible in the game

$$[N^t]a \wedge [N^t]b \wedge [N^t]c \wedge [N^t]d$$

and that there are no in nite plays allowed

$$[\emptyset^t]\top$$

Finally, we specify the target effectivity function of the strategic game:

$$[1^t](a \vee b) \wedge [1^t](c \vee d) \wedge [2^t](a \vee c) \wedge [2^t](b \vee d)$$

Let Δ be the set consisting of these 8 axioms. Then Δ is satis able by an extensive game without simultaneous moves iff E^α_G is realized by this game. Hence, Δ is not satis able by an extensive game without simultaneous moves. On the other hand, considering G' with s_4 being replaced by s_3 in G and substituting c for d in the axioms appropriately yielding the set Δ', the extensive game of gure 4 satis es Δ'.

As for the logical analogue of proposition 8, the preservation of local properties on the global level, observe that the four playability conditions for E^t can be translated into the logical language:

$$
\begin{array}{ll}
(\perp) & \neg[C^t]\perp \\
(\top) & [N]\top \to [C^t]\top \\
(\text{N}) & [N]\top \to (\neg[\emptyset^t]\neg\varphi \to [N^t]\varphi) \\
(\text{S}) & ([C_1^t]\varphi_1 \wedge [C_2^t]\varphi_2) \to [(C_1 \cup C_2)^t](\varphi_1 \wedge \varphi_2) \\
 & \text{where } C_1 \cap C_2 = \emptyset
\end{array}
$$

Note that the antecedent of axioms ($\top$) and (N) is necessary to handle terminal states where no coalition is effective for anything. All four axiom schemas are valid for extensive games without in nite plays. Similarly, the majorativity condition can be translated into

$$[N^t]\varphi \to [C^t]\varphi$$

where $|C| > \frac{1}{2}|N|$ and shown to be valid for majorative effectivity frames.

Metatheory

The concern of the present paper has been mainly semantic in nature. We introduced a class of models and discussed various well-known subclasses, also showing how these models can be used to model scenarios of multi-agent interaction and what new questions are raised. On the logical side, there are a number of meta-theoretic questions which should be mentioned. In (Pauly 2000), the modal base logic (i.e. Coalition Logic without $[C^*]\varphi$) has been studied from a more logical perspective. As has been suggested in the previous section, the playability conditions can be translated into our logical language yielding an axiomatization which can be shown to be complete for the class of extensive games with simultaneous moves. Work is in progress on extending this axiomatization to the logic with $[C^*]\varphi$. Furthermore, the satis ability problem for the modal base logic was shown to be PSPACE-complete, and hence it is just as complex as the normal modal logic K. In contrast, introducing an iteration construct such as $[C^*]\varphi$ usually increases the complexity of the satis ability problem. In the case of Propositional Dynamic Logic (Harel 1984; Kozen & Tiuryn 1990), the satis ability problem is exponential-time complete, and we conjecture that the same is true for coalition logic.

Related Work

The semantics used here to formalize multi-agent ability is based on minimal models with a neighborhood relation for each agent. For the single agent case, such models have been used in (Brown 1988) to study the logic of ability. This logic of ability is a very weak modal logic since properties such as $\Diamond(A \vee B) \to (\Diamond A \vee \Diamond B)$ fail. The example given to illustrate the failure of this principle refers to a deck of cards turned face down. Since the colors (red or black) are concealed, I am not able to draw a red card nor am I able to draw a black card, while I am able to draw a card which is either black or red. From our perspective, we interpret the situation as a game against nature, i.e. as a 2-player game where Nature chooses which card to give to me. The advantage of this approach is that it makes the roles of the players explicit, and hence one can point out that if the situation were in fact a 1-player game, the modal distribution principle would hold after all.

The approach taken here to formalize the ability of groups of agents differs somewhat from the existing literature on multi-agent ability. Among the earliest works, (Tennenholtz & Moses 1989) conceive of an agent as a set of nite state machines whose transitions model the agent's actions. As in an extensive game with simultaneous moves, the joint actions of the agents determine the new con guration of the system. Of central concern is the *cooperative goal achievement* (CGA) problem: Is there a run of the system in which all agents achieve their goal? It is argued that this problem is PSPACE-complete, which seems to correspond nicely to the complexity of the satis ability problem of the basic modal coalition logic, but the decision problems are quite different. Note also that our logical approach is more general in that it is not speci cally tailored to the CGA problem alone, but also allows to ask e.g. whether an in nite run can be forced by some group of agents.

While the work of Werner in (Werner 1990) is more directly related in its logical approach, his framework includes much more than just ability, covering also time, intentions, actions, knowledge. Given the more complex aims, his semantics is much more complicated than what is proposed here, and some of the fundamental issues which arise purely on the level of abilities are not investigated, e.g. the relationship between local and global ability, basic cooperative axioms such as superadditivity, etc.

Related to Werner's work, (Wooldridge & Fisher 1992) also takes a logical approach to multi-agent interaction which includes communication between agents. Their notion of goal achievement essentially corresponds to α-effectivity. Their logic is very expressive (including "at least" rst-order logic) and hence also much more complex then the rather simple system presented here. Some of their axioms however have direct analogues in coalition logic, e.g. one of their axioms states that bigger groups cannot achieve less (coalition-monotonicity) which they write as $\forall x \forall y((Can(x,\varphi) \wedge x \subseteq y) \rightarrow Can(y,\varphi))$ where x and y refer to groups of agents.

Conclusions and Future Work

Note that our general approach is different from the works cited: The aim is to provide a formal logical theory of ability in a multi-agent setting, without adding any other notions such as beliefs, intentions etc. which would complicate the picture. What is more important, we want a general model of ability, and this is what effectivity functions allow us to do. Effectivity in game-like situations (which is taken as basic in the other approaches) is only a special case which can be characterized by certain axioms (proposition 2), precisely the properties of group ability which characterize strategic games. This approach still allows us to model situations which would be beyond the scope e.g. of (Wooldridge & Fisher 1992) because they violate the coalition-monotonicity mentioned (think e.g. of RoboCup where a team of robots playing soccer may lose its ability to win if a completely malfunctioning robot is added to the team which always blocks the goal of the opposing team). Even this relatively simple model however is suf cient to ask many of the questions raised in the literature such as the CGA problem or the maintenance/achievement agent design problems of (Wooldridge 2000).

For the computer scientist, the work presented in this paper should provide an interesting generalization of work initiated in (Dijkstra 1976) on partial and total program correctness. Moving from programs to multi-player processes, one gets a better picture of the assumptions needed to establish various connections between partial and total correctness (proposition 1). For the game-theorist on the other hand, the distinction between partial and total effectivity provides the conceptual tool to analyze situations where certain coalitions have the power to force an impasse through in nite looping. Furthermore, we hope to have shown that effectivity frames are a useful model for dynamic processes and raise some new questions as well. The fact that a logic can be associated to these effectivity frames provides not only a conceptual link between game theory and logic, but also an algorithmic approach to solving game-theoretic questions.

Besides the open meta-theoretic questions mentioned, there are also game-theoretic questions which lend themselves for future work. The literature on effectivity functions knows various concepts of effectivity which differ from the notion of α-effectivity employed here and (implicitly) in most of the literature. The general approach adopted here equally well applies to these other notions such as β-effectivity, since global and terminal effectivity are not formally tied to any particular notion of effectivity, an advantage of the generality of our approach.

On a more applied note, examples such as the legislative procedure given in gure 2 suggest looking at more realistic cases of political/social processes where coalition formation is involved. As shown e.g. in (Vannucci 2000), effectivity functions can play a role in such an analysis, and it would be interesting to see how effectivity frames and Coalition Logic could be useful here. The domain of mutli-agent systems however is much broader than the narrow range of examples considered here, and investigating how much coalition logic can contribute to the study of more typical examples of multi-agent systems remains a task for the future.

References

Abdou, J., and Keiding, H. 1991. *Effectivity Functions in Social Choice*. Kluwer.

Abdou, J. 1998. Rectangularity and tightness: A normal form characterization of perfect information extensive game forms. *Mathematics of Operations Research* 23(3):553–567.

Binmore, K. 1992. *Fun and Games – A Text on Game Theory*. Heath.

Brown, M. 1988. On the logic of ability. *Journal of Philosophical Logic* 17:1–26.

Dijkstra, E. 1976. *A Discipline of Programming*. Prentice-Hall.

Harel, D. 1984. Dynamic logic. In Gabbay, D., and Guenthner, F., eds., *Handbook of Philosophical Logic*, volume II. Reidel.

Kozen, D., and Tiuryn, J. 1990. Logics of programs. In van Leeuwen, J., ed., *Handbook of Theoretical Computer Science*, volume B. MIT Press.

Moulin, H., and Peleg, B. 1982. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics* 10(1):115–145.

Moulin, H. 1983. *The Strategy of Social Choice*. North-Holland.

Osborne, M., and Rubinstein, A. 1994. *A Course in Game Theory*. MIT Press.

Parikh, R. 1985. The logic of games and its applications. In Karpinski, M., and van Leeuwen, J., eds., *Topics in the Theory of Computation*, volume 24 of *Annals of Discrete Mathematics*. Elsevier.

Pauly, M. 2000. A modal logic for coalitional power in games. *Journal of Logic and Computation*. To appear.

Peleg, B. 1998. Effectivity functions, game forms, games and rights. *Social Choice and Welfare* 15:67–80.

Tennenholtz, M., and Moses, Y. 1989. On cooperation in a multi-entity model (preliminary report). In *Proceedings of the Eleventh International Joint Conference on Arti cial Intelligence (IJCAI-89)*, 918–923.

Vannucci, S. 2000. Effectivity functions and parliamentary governance structures. In Patrone, F.; Garc´a-Jurado, I.; and Tijs, S., eds., *Game Practice: Contributions from applied game theory*. Kluwer. 225–239.

Werner, E. 1990. What can agents do together? In *Proceedings of the 9th European Conference on Arti cial Intelligence (ECAI-90)*, 694–701.

Wooldridge, M., and Fisher, M. 1992. A rst-order branching time logic of multi-agent systems. In *Proceedings of the 10th European Conference on Arti cial Intelligence (ECAI-92)*, 234–238.

Wooldridge, M. 2000. The computational complexity of agent design problems. In *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*.